DOCUMENT RESUME

ED 186 458 TM 800 148

AUTHOR	Erviti, Vivian; And Others
TITLE	The Development of Rating Scales to Assess the Performance of Medical Students in Clinical Settings.
INSTITUTION	National Board of Medical Examiners, Philadelphia, Pa.
SPONS AGENCY	Department of Health, Education, and Welfare, Washington, D.C.
PUB DATE	[Nov 79]
CONTRACT	231-76-0067
NOTE	18p.; Paper presented at the Annual Meeting of the American Association of Medical Colleges (Washington, DC, November, 1979).
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Clinical Experience; *Competence; Higher Education; Medical School Faculty; *Medical Students; Physicians; Pretesting; *Rating Scales; Self Evaluation (Individuals); Student Evaluation; *Test Construction; Test Reliability

ABSTRACT

A rating scale was developed for the evaluation of behaviors and skills required of medical students. From 400 items based on various role-defining studies, 92 were selected. The project staff randomly assigned 46 items to each of two alternate forms. Seven medical schools were identified as sites for pilot testing. Medical students who agreed to participate were asked to rate themselves and to agree to be rated by two or more resident and faculty supervisors. Ratings were distributed across the four response options available (seldom, sometimes, usually, consistently) on most of the 92 items. Raters were not reluctant to use the more negative responses at times. Very few students used the fifth option (don't know), although physicians used it rather freely. There was a significant inverse relationship between the frequency of use of the don't know option and the number of hours of contact with a student. Interrater reliability ranged from moderate to substantial agreement. The staff had hypothesized that residents would have more extended contact with students than faculty, be more knowledgeable about student performance, and therefore provide more reliable ratings. The data supported the first two hypotheses, but not the third. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186458

FEB 1980 RECEIVED ERIC/TM

THE DEVELOPMENT OF RATING SCALES TO ASSESS THE PERFORMANCE OF MEDICAL STUDENTS IN CLINICAL SETTINGS

Vivian Erviti, Ph.D., Lawrence J. Fabrey, M.S. and Jane V. Bunce
National Board of Medical Examiners

Introduction

Traditionally, evaluation has played an important role in a physician's educational experiences from the time of application for admission to medical school through the time when the physician is ready to assume independent responsibility for patient care. The assessment of physician competency is frequently based on the completion of a specified period of training, performance on written and/or oral examinations, ratings of traits and personality characteristics and ratings of clinical skills. Each of these methods can be effective in assessing some aspects of physician competency.[1,2,3,4]

The evaluation of the clinical performance of medical students is frequently based on ratings of behavior in clinical settings. A rating involves an evaluative summary of past or present experiences in which the rater processes data in a complex and individualistic way to arrive at a judgment. Many factors affect the reliability and validity of the ratings obtained. Raters may be unwilling to spend the time and energy necessary to provide accurate ratings or they may over-identify with the individual being rated and produce biased ratings. Other factors which interfere with the validity of the judgments include the lack of opportunity for the rater to observe the ratee, the covertness and/or ambiguity of the behavior, a lack of a uniform standard to use in interpretation of either the stimulus item or the response option and individual rater biases and idiosyncrasies.[5] Thus, in order to assure that a scale will produce reliable and valid data, it should be developed carefully, pilot tested thoroughly and revised until reliability is adequate and procedures for the selection of raters and for scale administration have been specified.

The study described in this report was designed to develop a rating scale for the reliable evaluation of some of the behaviors and clinical skills required of medical students. This tool may be used as a potential criterion of performance for other measurement instruments such as written examinations or as a measure of the competencies of medical students that cannot be assessed by other means.

Methodology

In order to begin the development of rating scales for the assessment of the competencies of physicians, a comprehensive list of the components of clinical competence was produced by reviewing the critical incidents study performed for the National Board of Medical Examiners (NBME) by the American Institutes for Research.[6] Nine general components of clinical competence were identified:

The research reported in this article was supported in part by the Department of Health, Education, and Welfare, Health Resources Administration, Contract Number 231-76-0067.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Vivian Erviti

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800 48

a. History Taking Competence
b. Physical Examination Skill
c. Test Technique and Understanding
d. Diagnostic Acumen
e. Wisdom in Deciding on Appropriate Care
f. Judgment and Skill in Implementing Care
g. Competence of Continuing Care
h. Effectiveness of Physician-Patient Relationship
i. Responsibility as a Physician

A project advisory committee of five physicians and one clinical psychologist was appointed and met several times to refine the definition of clinical competence by adding to and clarifying the list of components and then to use these components as a framework within which to develop behavioral statements which could serve as rating scale items. The committee was charged to write behavioral statements which would be applicable to physicians in all medical specialties and to include statements which describe a wide range of physician behavior from highly positive to highly negative. Approximately 400 items were generated.

The items were carefully reviewed by members of the project advisory committee and NBME staff to determine whether each item represented a major component of clinical competence and whether medical students might be expected to vary on the item. In addition, a questionnaire was mailed to faculty in four medical schools in the Philadelphia area to determine whether faculty members considered the items appropriate for rating the clinical performance of medical students. Two faculty members in each of the required clerkships were asked to indicate which behavioral statements were appropriate and 19 out of 40 questionnaires were returned.

Using all data obtained from the surveys and staff reviews, the committee selected 92 items for inclusion in the scales. The items were selected proportionately from each of the nine general components of competence. All items selected were considered appropriate for students in basic clerkships and those in their most advanced level of a clerkship in primary care disciplines. These students may be in their second, third, or fourth year of medical school, depending upon the program of the student and the curriculum of the participating school. Subsequent to the selection of the 92 items, the project staff decided to create two forms of the rating scale instrument. Each of the two forms contained 46 items that were randomly selected, within the nine general competency categories, from the total 92-item pool. Some sample items are shown in Table 1, along with the five response options provided for each item.

Seven medical schools were identified as sites for pilot testing. The NBME staff first contacted the dean's office at each participating institution to obtain the necessary introduction to departments of family practice, internal medicine, obstetrics and gynecology, pediatrics, and surgery. Administrators in the departments agreeing to participate were provided with packets of materials to be used in soliciting student participants. Each medical student who agreed to participate was asked to rate himself/herself and to agree to be rated by two or more supervisors. Each departmental administrator identified those residents and faculty members who had intimate knowledge of a participating student's performance and asked them to rate the student. The administrator was asked to identify at least two resident/faculty raters if it was possible to do so. All participants from a given institution used the same form of the instrument.

Results

A total of 263 completed rating scales were returned to the NBME. Data contained in 26 of the rating scales were eliminated from the analyses. Twenty scales were eliminated because they contained at least one-half of the 46 item responses in the Don't Know category or were left blank. Although it was useful to know that one resident and 11 faculty members were unable to provide responses for a large number of items, it was determined that use of these data could be misleading. Four other ratings were considered invalid because one physician returned identical rating scales for four different students using photo-duplicated copies. Another rating considered invalid was received from a student who provided no variation in response; that is, he responded Usually to every item, regardless of whether the item was positively or negatively worded. The last rating eliminated was the rating of a student who had inadvertently used Form B in an institution that was assigned Form A.

Of the 237 rating scales remaining in the final sample to be used for the analyses, 135 were Form A and 102 were Form B. Table 2 shows the number of students submitting usable data as well as the number of rating forms submitted by residents and faculty members at each of the seven participating institutions. From the five schools using Form A, usable rating scales were received from 64 students, while usable Form B scales were received from 47 students. Thirteen different residents and 31 faculty members submitted 20 and 51 usable Form A ratings, respectively. For Form B, 14 residents and eight faculty members provided 32 and 23 usable ratings, respectively.

As shown in Table 3, 40 ratings were received from students who had no corresponding rating from a supervisor. Fifty-two students had one physician provide a rating, 17 had two physician raters, ten had three other raters and only two students had four supervisor ratings of their performance in addition to their own. The total number of students shown in Table 3 exceeds the 111 individuals mentioned previously because the figure in Table 3 includes 11 physician ratings of eight students who returned a completed consent form but not a self-rating and two students who completed two ratings each in different clerkships.

As mentioned previously, 44 physicians returned Form A ratings and 22 used Form B. Table 4 shows that half of the physicians returned ratings on only one student, while slightly more than 40% returned ratings on either two or three students. Six percent of the physicians returned ratings on four students and one physician (approx. 2%) rated eight students.

In general, students were able to provide a rating on more of the 46 items than physicians. Fifty-one percent of the students did not use the Don't Know option for a single item and 91% used it less than five times while completing the scale. However, only 58% of the residents and 23% of the faculty used it less than five times. When the total sample of 263 rating scales was considered, the Don't Know response was chosen as many as 27 times by an individual resident and as many as 39 times by a faculty rater. Even following the elimination of the forms with an excessive number of Don't Know ratings (predominantly faculty), the mean number of times this option was used was significantly different for the two rater groups ($t=3.90$, $p < .001$) as shown in Table 5. The mean number of hours of contact with a student reported by residents and faculty was also significantly different ($t=8.77$, $p < .001$) as shown in Table 5. In addition,

there was a significant inverse relationship ($r = -.52$, $p < .001$) between these two variables; that is, physicians who report spending relatively more time with students tend to use the Don't Know response less.

The distribution of ratings across the four response options for each item were tabulated for students, residents and staff physicians separately and the means and variances of the items calculated. (See Table 1 for distributions of physicians' ratings across options for ten sample items.) Most items showed a good distribution across the options and concentration on one or two of the options occurred rarely. None of the items on Form A showed a concentration of responses on a single response option and only three items on Form B had such a concentration. On Form A, student self-ratings covered three or more options for 39 of the 46 items. Residents used more than two options for 33 items and faculty used three or more on 43 items. The 46 items on Form A had a mean of 3.27 and a standard deviation of .66 on the four-point scale from Seldom to Consistently, with four indicating the most positive behavior. Individual item means ranged from 1.99 to 3.94, and standard deviations ranged from .22 to 1.06. For Form B, students covered three or more options on 43 items, residents covered three or more options on 39 items and staff physicians used them on 34 items. The mean of the item means for Form B was 3.29, ranging from 2.50 to 3.97. The mean standard deviation was .67 with a range of .24 to .97. Thus, all response options were used by some raters at some times for most of the items and the range of item means indicates that raters were not reluctant to use the more negative response options at times.

An important question to be answered in this pilot test of rating scales for medical students concerned the reliability of the data. Although there are a variety of methods to estimate interobserver agreement (e.g., simple proportion of agreement or various correlation coefficients) the weighted Kappa coefficient was chosen as the most appropriate method for the rating scale developed in this study. In general, Kappa statistics are useful for qualitative data and are calculated as a proportion of agreement with a correction for chance agreement. Weighted Kappa is a variation of the Kappa statistic which penalizes totally discordant responses but provides proportionally more credit as the responses of two raters approach perfect agreement. Weighted Kappa should be used when responses are not strictly nominal, particularly when there is reason to believe that differences between response options are of different quantitative value.[7] Using the example presented in Cohen,[8] making a mistake in the clinical classification of a neurotic as a psychotic is more serious than classifying a neurotic as having a personality disorder. The project staff believed that the anchors (Seldom, Sometimes, Usually, and Consistently) used on the rating scale to evaluate medical student performance presented an analogous situation; that is, there would appear to be differential intervals between the four scale points. Therefore, weighted Kappa was more appropriate than unweighted Kappa in this study.

Some authors[8,9] have provided suggestions for determining appropriate weights for calculating weighted Kappa and these typically rely on purely common sense decisions. An informal scaling study, based partially on a study by Spector,[10] was performed to provide a basis for assigning disagreement weights. Approximately 50 professional staff from the NBME were asked to complete a survey instrument to determine how people perceive the four anchors used on the rating scale, that is, to attempt to quantify these qualitative terms in order to assign disagreement weights.

Thirty-seven individuals completed the survey instrument. These data provided support for the staff's hypothesis of a perceived neutral point and a relatively larger interval between the responses Sometimes and Usually than between Seldom and Sometimes, or between Usually and Consistently. If roughly equal intervals had been estimated by the survey respondents (e.g., 20, 40, 60, 80%), strictly linear weighting would have been implemented. However, it was obvious that these words are not perceived to be equally spaced, rather there exists a large gap between the two middle points on the scale.

The project staff evaluated the data and considered the severity of various rating errors in order to assign the disagreement weights shown in Table 6. The actual disagreement weights are unimportant, rather the relative (or proportional) values are critical to calculation of weighted Kappa. Table 6 shows that the most serious discrepancy between two raters occurs when one responds Seldom and the other Consistently. Similarly, it can be seen that an error between Seldom and Usually can be considered five times as serious as an error between Usually and Consistently.

Weighted Kappa statistics were calculated for each pair of raters evaluating a given medical student. Kappa coefficients were computed for a total of 173 rater pairs. These included all types of rater pairs (resident-resident, faculty-faculty, resident-faculty, resident-student and faculty-student) on each form of the instrument. A frequency distribution of the resulting Kappa statistics was developed and is shown in Table 7. There were no major differences in level of agreement between the three types of possible physician pairs; that is, two residents, two faculty members or one resident and one faculty member rating a given student. The Kappa statistics shown in Table 7 indicate that most physicians demonstrate at least moderate[11] agreement about the aspects of clinical competence assessed by the specific rating scale statements. In other words, it could be said that both Form A and Form B can provide reasonably reliable descriptions of student behavior. Another aspect of the reliability of the instrument is demonstrated by the last two columns for each form shown in Table 7. These columns indicate levels of agreement between a physician's and a student's self rating that are similar to those observed for pairs of physicians. Overall, 41% of the 173 rater pairs demonstrated at least substantial agreement (greater than or equal to .61) and 77% were at least at the moderate agreement level (.41 or higher). Less than 7% of these rater pairs indicated less than fair agreement (.20 or less).

In addition to established reliability, a rating scale requires some assurance of validity if it is to have utility. No satisfactory external criterion existed that could be used as a concurrent measure of clinical competence in this study. Therefore, the assurance of the validity of this rating scale is dependent upon its careful development as described in an earlier section of this report. Items were originally written and have subsequently been selected to reflect many aspects of the clinical performance of medical students. In addition to the input of the staff, a variety of experts were consulted to assist in the selection of items and the development of the scale.

Although less desirable than careful development to assure content validity, the scale appears to possess a fair degree of face validity. All contacts at the participating schools, including associate deans and education officers of specific departments, were eager to assist the NBME in the development of this rating instrument. The only reservation by any individual participating was the

length of the scale and the possibility of time conflicts with existing evaluation procedures. However, individuals expressing this concern typically stated that the rating scale used in this pilot test was superior to any existing evaluation form because it covered more completely the range of behaviors expected of medical students.

All individuals participating as raters were asked to respond to a user survey. Raters were generally moderately satisfied that the instrument could provide a comprehensive description of the performance of a medical student. Table 8 shows that 78% (171/219) of the raters responding indicated that they were moderately satisfied with the scale, although only 2% (5/219) were completely satisfied with the scale. Forty-three raters (20%) reported no satisfaction with the scale. Several of these unsatisfied individuals provided no further comment or reason for their dissatisfaction. Numerous participants commented that the items were too subjective and that there were too many of them. These individuals suggested that the number of items could be reduced by using general characteristics such as motivation, intelligence and ability to work with others as stems. Several faculty members stated that the amount of faculty-student interaction was inadequate to allow them to use the items intelligently.

Discussion

Reliable and valid rating scales are difficult to construct because of the many variables that can affect the quality of the data. Raters differ in the leniency of their ratings and also in their willingness to use all available response options. Interpretation of the items and response options are frequently highly subjective and may vary widely from one rater to another. Although individual differences in ratings were noted in the data collected from the seven institutions, all response options were used for most items. Therefore, the items appear to discriminate appropriately among students.

The purpose for preparing two forms of the scale was to reduce the amount of time required of any participant and still collect data on all 92 items. Because no comparison between the two forms was intended, the sample of students being rated on each form was neither random nor matched in any way. Differences between the means and standard deviations for the two forms may be an artifact of the particular cluster of items, the result of the different samples of students being rated, or resident and faculty differences.

The staff had hypothesized that residents would have more extended contact with students than faculty, be more knowledgeable about student performance and therefore provide more reliable ratings. The data support the first two hypotheses as evidenced by significant differences in the number of contact hours reported, the number of Don't Know responses, and the significant inverse relationship between these two variables. However, the third hypothesis concerning increased reliability of resident ratings was not supported.

Conclusions and Implications

The development of any evaluative tool is an iterative process. The data obtained in an initial pilot trial can be used to indicate modifications that must then be tested in further trials. The use of multiple data sources and multiple field trials is especially important in developing rating scales because of the large number of sources of error involved in ratings. A review of the Kappa coefficients for the individual rating scale items will enable the project staff

to select clusters of items that may improve the overall reliability. It is possible that the elimination of some items may increase reliability. Also, a two-point scale may be more appropriate based on the marked improvement in the Kappa coefficient when two rather than four response options were used.

In addition to the use of reliability coefficients, item elimination may be based on the failure of an item to discriminate among students, the fact that a large number of residents and faculty reported they didn't know about the specified aspect of behavior, the redundancy of items and an editorial review using generally accepted guidelines for rating scale construction.

The data collected in this initial pilot test of the 92 rating scale items will form the beginning of a data pool on individual items. It is possible that in the future the scale may be revised on the basis of the initial data available and subjected to additional pilot testing. A refined item pool could be used to customize scales for various groups of raters, clerkships, institutions, etc.

References

1. Randolph, J.G., Altman, R.P. & Anderson, K.D. Experience with in-service examinations of pediatric surgical trainees. Journal of Pediatric Surgery, December 1977, 12(6), 857-860.

2. Gallagher, R., Donnelly, M., Scalzi, P.M. & Deighton, M. Toward a comprehensive methodology for resident evaluation. Proceedings of the Sixteenth Annual Conference on Research in Medical Education, Washington, D.C., November 9-10, 1977.

3. Tudor, J.M. Performance documentation: How to confirm a resident's progress. Journal of Medical Education, April 1978, 53, 337-343.

4. Stillman, P.L. & Sabers, D.L. Using a competency-based program to assess interviewing skills of pediatric house staff. Journal of Medical Education, June 1978, 53, 493-496.

5. Thorndike, R.L. & Hagen, E.P. Measurement and evaluation in psychology and education (4th ed.). New York: John Wiley & Sons, 1977.

6. The Definition of Clinical Competence in Medicine: Performance Dimensions and Rationales for Clinical Skill Areas. A report of research conducted for the National Board of Medical Examiners by the American Institutes for Research, Palo Alto, CA., April 1960.

7. Fleiss, J.L. & Cohen, J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement, 1973, 33, 613-619.

8. Cohen, J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bulletin, October 1968, 70(4), 213-220.

9. Cicchetti, D.V. Assessing inter-rater reliability for rating scales: Resolving some basic issues. British Journal of Psychiatry, 1976, 129, 452-456.

10. Spector, P.E. Choosing response categories for summated rating scales. Journal of Applied Psychology, 1976, 61(3), 374-375.

11. Landis, J.R. & Koch, G.G. The measurement of observer agreement for categorical data. Biometrics, March 1977, 33, 159-174.

Table 1

Sample of Rating Scale Items and Physician Responses

		Seldom	Sometimes	Usually	Consistently	Don't Know
1.	Reevaluates abnormal laboratory values which are inconsistent with the clinical picture	0	11	28	12	3
2.	Obtains additional data to confirm impressions	1	8	29	13	3
3.	Integrates clinical data into one or more pertinent diagnoses	2	8	28	16	0
4.	Tends to be indecisive, frightened, or timid when participating in emergency therapy	5	2	1	1	45
5.	Ignores the advice and downgrades the roles of the other members of the health care team	44	1	0	0	9
6.	Facilitates accurate information gathering by establishing rapport	0	9	31	26	5
7.	Interprets abnormal test results properly	5	7	47	12	0
8.	Identifies relevant information from the mass of data gathered for each patient	1	9	44	17	0
9.	Is careless in performing procedures	45	2	0	0	14
10.	Attempts procedures with improper equipment and/or inadequate knowledge or skills	47	2	0	5	22

Table 2

Number of Rating Forms From Students, Residents and Faculty in Final Sample

Form	School No.	# Rating Forms from Students	# Rating Forms from Residents	# Rating Forms from Faculty
A	1	8	7	12
	2	13	6	10
	4	3	3	1
	6	37	2	27
	7	3	2	1
B	3	23	7	3
	5	24	25	20
	TOTAL NO. FORMS:	111	52	74

Table 3

Number of Students for Whom One or More Ratings Were Obtained

Number of Ratings	Number of Students Form A	Number of Students Form B
Only Self Rating	24	16
1 Resident or Faculty Rating	38	14
2 Resident or Faculty Ratings	1	16
3 Resident or Faculty Ratings	9	1
4 Resident or Faculty Ratings	1	1
Total	73 (135 Ratings)	48 (102 Ratings)

Table 4

Number of Physicians Providing One or More Ratings of Students

No. of Ratings	Form A		Form B	
	No. of Residents	No. of Faculty	No. of Residents	No. of Faculty
1	9	18	4	3
2	1	8	3	1
3	3	3	6	2
4	0	2	1	1
8	0	0	0	1
	13	31	14	8

Table 5

Comparison of Resident and Faculty Raters

Raters	No. of Don't Know Responses $\bar{X}$	S.D.	No. Hours Contact with Student $\bar{X}$	S.D.
Residents	5.10	5.48	37.92	18.09
Faculty	9.31	6.30	14.38	11.94
	$t = 3.90$ $(p < .001)$		$t = 8.77$ $(p < .001)$	

$r = -.52$ $(p < .001)$

Table 6

Disagreement Weights and Summary of Responses to Informal Scaling Survey

Anchor	Disagreement Weights				Scaling Survey Statistics	
	Seldom	Sometimes	Usually	Consistently	Mode, %	Mean, %
Seldom	0	2.0	7.5	9.0	5%	6.7%
Sometimes	2.0	0	5.5	7.0	25%	31.5%
Usually	7.5	5.5	0	1.5	80%	77.4%
Consistently	9.0	7.0	1.5	0	90,95, 99,100%	96.5%

Table 7

Agreement Between Rater Pairs Rating the Same Medical Student on 46-Item Rating Scales

Weighted Kappa	Level of Agreement	Number of Weighted Kappa Coefficients in Each Range									
		Form A					Form B				
		Res-Res	Fac-Fac	Res-Fac	Res-Stu	Fac-Stu	Res-Res	Fac-Fac	Res-Fac	Res-Stu	Fac-Stu
< 0	Poor	0	2	0	0	1	0	0	0	0	1
0-.20	Slight	1	0	0	0	4	0	2	0	0	1
.21-.40	Fair	0	1	2	4	6	4	0	0	6	5
.41-.60	Moderate	1	6	4	5	17	4	2	5	11	7
.61-.80	Substantial	3	2	8	6	14	2	2	2	12	9
.81-1.0	Almost Perfect	0	1	3	1	2	0	2	1	0	1
Total No. of Pairs		5	12	17	16	44	10	8	8	29	24
Average Weighted Kappa		.5495	.4524	.6313	.5648	.5193	.4619	.5086	.6000	.5371	.5148
Standard Deviation		.253	.302	.165	.199	.226	.121	.323	.137	.166	.230

Table 8

Satisfaction with Rating Scale: Summary of Responses to Question 3 on the User Survey

How satisfied are you that this instrument can provide a comprehensive description of the clinical performance of a medical student in a basic or more advanced but general clerkship?

Type of Rater	Number of Participants Responding: Not at all satisfied 1	Moderately satisfied 2	3	Completely satisfied 4	No response	Mean Satisfaction Rating
Form A:						
Students	7	30	22	0	5	2.25
Faculty	2	7	8	2	1	2.53
Residents	13	12	21	1	4	2.21
Form B:						
Students	8	16	18	2	4	2.32
Faculty	3	13	13	0	3	2.34
Residents	10	7	4	0	2	1.71
Overall	43	85	86	5	19	2.22